

# THE PUZZLE OF “EX ANTE EFFICIENCY”: DOES RATIONAL APPROVABILITY HAVE MORAL WEIGHT?

MATTHEW D. ADLER<sup>†</sup>

## INTRODUCTION

What is the moral status of ex ante efficiency? Many within the law and economics tradition seem to think that the ex ante efficiency of legal institutions has moral weight—that legal officials have a moral reason to create ex ante efficient institutions.<sup>1</sup> I think the view is questionable.

---

<sup>†</sup> Professor, University of Pennsylvania Law School. Many thanks to Daniel Farber, Eric Orts, the participants in the University of Pennsylvania Law School Symposium on Preferences and Rational Choice, and my interlocutors at an ad hoc Penn Law workshop for their comments. Special thanks to Stephen Perry for his detailed, written comments.

<sup>1</sup> A recent and visible attempt by two prominent law and economics scholars to specify the discipline’s appropriate methodology explicitly defends, and at many junctures deploys, the criterion of ex ante efficiency. See LOUIS KAPLOW & STEVEN SHAVELL, FAIRNESS VERSUS WELFARE 437-43 (2002) (arguing for the ex ante perspective). A full assessment of the role that ex ante efficiency, in one sense or another, has played within law and economics is well beyond the scope of this Article. It seems clear, however, that this concept (in various specific formulations) has been quite important for this scholarly tradition. For recent examples of law and economics work that, in evaluating legal doctrines or other legal institutions, explicitly invokes the criterion of “ex ante efficiency” or “ex ante welfare maximization” (a criterion that is, again, specifiable in various ways), see Stephen J. Choi & Eric L. Talley, *Playing Favorites with Shareholders*, 75 S. CAL. L. REV. 271 (2002); Robert Cooter & Ariel Porat, *Anti-Insurance*, 31 J. LEGAL STUD. 203 (2002); Lucian Arye Bebchuck, *Property Rights and Liability Rules: The Ex Ante View of the Cathedral*, 100 MICH. L. REV. 601 (2001); Barry Adler et al., *Regulating Consumer Bankruptcy: A Theoretical Inquiry*, 29 J. LEGAL STUD. 585 (2000); Yongmin Chen, *Promises, Trust, and Contracts*, 16 J.L. ECON. & ORG. 209 (2000); Ian Ayres & Kristin Madison, *Threatening Inefficient Performance of Injunctions and Contracts*, 148 U. PA. L. REV. 45 (1999); Lucian Arye Bebchuck & Andrew T. Guzman, *An Economic Analysis of Transnational Bankruptcies*, 42 J.L. & ECON. 775 (1999); Yeon-Koo Che & Alan Schwartz, *Section 365, Mandatory Bankruptcy Rules and Inefficient Continuance*, 15 J.L. ECON. & ORG. 441 (1999); Omri Ben-Shahar, *Should Products Liability Be Based on Hindsight?*, 14 J.L. ECON. & ORG. 325 (1998). It also bears noting that Richard Posner’s (in)famous defense of “wealth maximization” invoked an ex ante perspective by arguing that judicial use of the wealth-maximization norm would, hypothetically, be consented to by (almost) all actors choosing under conditions of uncertainty. See RICHARD A. POSNER, *THE ECONOMICS OF JUSTICE* 88-115 (1981) (exploring the “ethical and political basis of wealth maximization”).

In this Article, I seek to undermine the view by attacking a more general claim; the putative moral weight of ex ante efficiency is a particular variant of this generic claim, one that marries the claim with a particular view about the nature of rationality. The generic claim is that *rational approvability* has moral weight—that legal officials have moral reason to choose actions that citizens would be rational to approve.

Part I fleshes out the concept of rational approvability and describes the link between rational approvability and ex ante efficiency. Part II asks whether rational approvability has moral force given the particular account of rationality dominant within economics: subjective expected utility (SEU). I consider various arguments for an affirmative answer—grounded in welfare, agency, stability, and the legal official's own uncertainty—and reject them all. Part III assesses the moral force of rational approvability on the assumption that SEU has been replaced with alternative accounts of rationality. What is rational for someone to do or approve might be a matter of maximizing objective expected utility, realized utility, subjective expected value, objective expected value, or realized value. I suggest that these variations on SEU either fail to be plausible accounts of rationality or, if plausible, generally fail to support the claim that rational approvability has moral weight.

---

Besides the Kaplow and Shavell opus, two other important recent attempts to defend the legal relevance of ex ante efficiency are Daniel A. Farber, *Economic Efficiency and the Ex Ante Perspective*, in THE JURISPRUDENTIAL FOUNDATIONS OF CORPORATE AND COMMERCIAL LAW 54 (Jody S. Kraus & Steven D. Walt eds., 2000); Christopher T. Wonnell, *Efficiency and Conservatism*, 80 NEB. L. REV. 643 (2001). I do not discuss the Wonnell, Farber, or Kaplow and Shavell defenses of ex ante efficiency here, in part because I am not sure that they mean to defend the version of ex ante efficiency I am criticizing. My variant strictly incorporates the subjective expected utility (SEU) account of rationality standard within economics; the authors just mentioned might not give weight to individuals' actual probabilistic beliefs if mistaken. See KAPLOW & SHAVELL, *supra*, at 1330-34 (recognizing that individuals have imperfect information and suggesting that, if the government has superior information, legal decisions should reflect that); Wonnell, *supra*, at 682 n.105 (stating that the state's probability estimates should govern where more accurate than the individual's, unless the latter estimates are tied to personal identity). But see KAPLOW & SHAVELL, *supra*, at 441 ("If . . . individuals would unanimously prefer regime A to regime B if there were a referendum, we do not understand how it can be argued that, on occasions when no vote is possible and thus the government officials themselves must decide between A and B, they may pick B." (footnote omitted)).

The relevance of ex ante efficiency is, of course, a problem for welfare economics generally, not just law and economics. For an overview of the literature, see Philippe Mongin & Claude d'Aspremont, *Utility Theory and Ethics*, in 1 HANDBOOK OF UTILITY THEORY 371, 437-44 (Salvador Barberà et al. eds., 1998).

## I. EX ANTE EFFICIENCY AND RATIONALITY

SEU is the standard account of rationality within economics,<sup>2</sup> and stipulates the following, given some actor in a choice situation at time  $T$ , choosing among actions  $A_1 \dots A_n$ : (1) each choice,  $A_i$ , can be represented as a lottery over possible worlds, where the probabilities in the lottery are numerical measures of the actor's subjective degrees-of-belief that the choice, if made, will result in the various possible outcomes; (2) the actor's preferences over possible worlds can be represented by a numerical "utility" function; (3) for each choice, an *expected utility* can be calculated by using the subjective probability numbers in the matching lottery to discount the utilities of the various possible outcomes, and then aggregating; and (4) the rational choice is the choice with the largest expected utility.

It is common in law and economics scholarship to describe some law, doctrine, rule, statute, or other legal institution as maximizing everyone's expected utility and therefore "ex ante efficient." Note the problems in this formulation, at least within an SEU account of expected utility. First, subjective expected utility, strictly construed, is a

---

<sup>2</sup> Good introductions to the literature on rational choice, with a focus on SEU and other such accounts structured in terms of the actual or possible outcomes resulting from choice, include ELLERY EELLS, *RATIONAL DECISION AND CAUSALITY* (1982); RICHARD A. FUMERTON, *REASON AND MORALITY: A DEFENSE OF THE EGOCENTRIC PERSPECTIVE* (1990); SIMON FRENCH, *DECISION THEORY: AN INTRODUCTION TO THE MATHEMATICS OF RATIONALITY* (1988); MICHAEL D. RESNIK, *CHOICES: AN INTRODUCTION TO DECISION THEORY* (1987). I take the term SEU from Eells. See EELLS, *supra*, at 4-86 (presenting and discussing the SEU account). The classic presentations of SEU are Frank Plumpton Ramsey, *Truth and Probability*, in *THE FOUNDATIONS OF MATHEMATICS AND OTHER LOGICAL ESSAYS* 156, 156-98 (R.B. Braithwaite ed., 1931); LEONARD J. SAVAGE, *THE FOUNDATIONS OF STATISTICS* (1954); R.C. JEFFREY, *THE LOGIC OF DECISION* (1965).

For support of the claim that SEU is the standard economic account of rational choice, see JACK HIRSHLEIFER & JOHN G. RILEY, *THE ANALYTICS OF UNCERTAINTY AND INFORMATION* 10 (1992) ("[W]e disregard [Frank] Knight's distinction [between risk and uncertainty], which has proved to be a sterile one. For our purposes risk and uncertainty mean the same thing. . . . [W]e will be dealing throughout with a 'subjective' probability . . . : probability is simply *degree of belief*."). Perhaps the claim is too strong. A weaker claim would be that the standard economic view of rationality is an expected utility view, with either subjective or objective probabilities (i.e., either SEU or objective expected utility (OEU)). See DAVID M. KREPS, *A COURSE IN MICROECONOMIC THEORY* 112 (1990) ("The von Neumann-Morgenstern model, where probabilities are objective, and the Savage . . . model, where probabilities are subjective, are the chief models of consumer choice under uncertainty in microeconomics."). I argue below against the moral force of rational approvability given OEU. *Infra* Part III.A. These arguments would undermine the claim that ex ante efficiency, redefined in terms of objective rather than subjective expected utilities, has moral weight.

feature of each person's choices. What does it mean to speak of the expected utility, for a person, of some other kind of thing—here, a legal institution? Second, subjective expected utility varies over time. A person's subjective expected utility for her own choices (and presumably for derivative things, such as institutions) depends upon her subjective probability assessments, which are not temporally fixed. A person's probability assessment at some time,  $T_1$ , that some action will produce some possible world may be different from her assessment at another time,  $T_2$ , that the action will produce that world.<sup>3</sup> Legal institutions, however, endure over time. It is not as if a legal institution exists only at one moment and we can (somehow) use everyone's subjective probability numbers at that time to determine whether the institution is *ex ante* efficient. So what time *is* relevant in determining *ex ante* efficiency?

Here is an attempt to resolve the problems and provide a clearer definition of *ex ante* efficiency. As just mentioned, rational choice theory is, centrally, prescriptive *for an actor* with regard to *the actor's own choices*. That is, the theory identifies the action, from the various possible actions open to person *P* in a choice situation, which *P* would be rational to choose. *P* rationally orders the possible actions of *P*. By extension, however, one can describe the rational ordering, *by some other person*, of *P*'s possible actions. Imagine an Agent and a Patient.<sup>4</sup> The Agent is in an actual choice situation. At a particular point in time,  $T$ , he is choosing among actions  $\{A_1 \dots A_n\}$ . Since these are the Agent's choices, not the Patient's, how does one construct a rational ordering of the choices by the Patient? I suggest that one might ask which choice by the Agent it would be rational for the Patient *to approve*. More precisely, we can imagine the Patient in a hypothetical choice situation at  $T$ , faced with the options  $\{A_1' \dots A_n'\}$ , where each  $A_i'$  is a determinative act of approval by the Patient: if the Patient were to choose  $A_1'$ , the Agent would do  $A_1$ ; if the Patient were to choose  $A_2'$ , the Agent would do  $A_2$ ; and so on. Since, at a minimum, any theory of rational choice can rank choices, any theory will be able to rank the Patient's hypothetical choices of approving one or another of the various choices open to the Agent.

---

<sup>3</sup> See *infra* notes 14, 36 (discussing the nature of probability, including subjective probability, and specifically discussing temporal variation in subjective probabilities).

<sup>4</sup> I use the term "Patient" because he is passive in this context—the person to whom choices are presented at  $T$  is the Agent, not the Patient—and because his interests or welfare may be affected by the Agent's action. The Patient is someone whom the Agent may need, morally, to take account of.

With this concept of rational approvability in hand, we can provide the following definition of ex ante efficiency, one that treats ex ante efficiency (as well as ex post efficiency) as a feature of particular, institution-affecting choices by legal officials:

*Ex Ante and Ex Post Efficiency:* Imagine that a legal official, the "Agent," is choosing among  $\{A_1 \dots A_n\}$  at some time  $T$ . The Agent's choices may affect legal institutions in various ways; they may create, support, destroy, or undermine such institutions. Some action  $A^*$  by the Agent is "ex ante efficient" if, for each individual in the population (each "Patient"), it maximizes the Patient's subjective expected utility at  $T$  to approve  $A^*$ , rather than the other actions in  $\{A_1 \dots A_n\}$ . Some action  $A_+$  by the Agent is "ex post efficient" if the actual world that would result, were  $A_+$  to be chosen, is preferred by all Patients at  $T$  to the actual worlds that would result, were the Agent to choose other actions from  $\{A_1 \dots A_n\}$ .<sup>5</sup>

This definition resolves the problems mentioned earlier. Given some institution-affecting choice by an Agent at time  $T$  and some group of "Patients" affected by the choice, for each individual Patient one can coherently assign subjective expected utilities to the Agent's choices and determine which choice it would maximize subjective expected utility for that Patient, at  $T$ , to approve. If the maximizing choice by the Agent, for all Patients, is the same choice, that choice is "ex ante efficient."

The problem of determining ex ante efficiency, given the temporal variability in any person's subjective probability assignments, is also solved by thinking of ex ante efficiency as a feature of particular institution-affecting choices by legal Agents. Choice situations presented to Agents, unlike whole legal institutions, can be thought of as occurring at particular times; it is the affected persons' subjective probabilities *at the time of the Agent's choice* that, I suggest, might be used to calculate the expected utilities that are determinative of ex ante efficiency.

Finally, the definition offered here draws a clean distinction between ex ante and ex post efficiency. It is common for legal scholars to draw this distinction. "Ex ante" and "ex post" efficiency are seen as

---

<sup>5</sup> The kind of efficiency defined here is, of course, Pareto superiority. Ex ante Pareto superiority might mean that everyone ex ante prefers the choice, or that at least one person does and no one prefers another choice. Similarly, ex post Pareto superiority might mean that everyone prefers the world that would result from the choice, or that at least one person does and no one prefers another world. See Amartya Sen, *Social Choice Theory*, in 3 HANDBOOK OF MATHEMATICAL ECONOMICS 1073, 1075 (Kenneth J. Arrow & Michael D. Intriligator eds., 1986) (distinguishing between "weak" and "strong" versions of the Pareto principle). The difference between these variants of ex ante and ex post efficiency is not relevant to my analysis in this Article, and the definition offered in the text is meant to be ambiguous between the variants.

logically distinct properties of legal institutions: some institution might be *ex ante* efficient but not *ex post* efficient, or vice versa. My definition warrants this sort of talk. Given some Agent placed in a choice situation at time  $T$ , choosing among  $\{A_1 \dots A_n\}$ , the choice that some Patient “*ex ante*” prefers, versus the choice that she “*ex post*” prefers—the choice that it maximizes subjective expected utility for her to approve, versus the choice whose outcome she most prefers—clearly can be different. Thus, too, it might clearly occur that the *ex ante* efficient choice by some Agent is not *ex post* efficient.

*Ex ante* and *ex post* efficiency, as I have clarified these concepts, cannot be ascribed to institutions simpliciter. Rather, they are ascribable to particular choices by legal officials: to possible actions by them, in particular choice situations at particular times. Is this a flaw in my conceptualization? I don’t think so. If we are to preserve the standard link between *ex ante* efficiency and the SEU account of rationality, I see no way to license talk of the *ex ante* efficiency of an institution simpliciter, detached from a particular time. Again, subjective probability assessments vary over time. And since SEU is fundamentally an account of rational *choice*, defining *ex ante* efficiency as, first and foremost, a feature of choices—and derivatively, of legal institutions, relative to particular choices that affect the institutions—seems the natural way to proceed. At a minimum, the choice-based account is a plausible conceptualization of *ex ante* efficiency and is the one I shall rely upon in this Article.<sup>6</sup>

---

<sup>6</sup> What alternative accounts are there? Consider these possibilities: an institution is *ex ante* efficient if (1) everyone prefers the world in which the institution exists to the world in which the institution does not exist; (2) everyone prefers the “nearest possible world” (to the actual world) in which the institution exists to the nearest possible world in which the institution does not exist; (3) everyone prefers a lottery over the possible worlds in which the institution exists to a lottery over the possible worlds in which the institution does not exist; or (4) everyone, at the time of the creation of the institution, prefers a lottery over the possible worlds in which the institution exists to a lottery over the possible worlds in which the institution does not exist.

The problems with (1) are that probability assessments and expected utility do not come into play here and further that institutions exist, and fail to exist, in multiple worlds; unless a person prefers all the worlds in which the institution exists to all the worlds in which the institution fails to exist, her preference will be undefined. The problem with (2) is that this, like (1), provides a conception of *ex post*, not *ex ante*, efficiency; as with (1), the preferences here are preferences over worlds represented by utility numbers, not preferences over lotteries represented by expected utilities calculated using subjective probabilities. The problem with (3) is that there are multiple times at which persons might be making the expected utility calculations over the lotteries; unless a person’s expected utility at all times for the institution lottery is greater than her expected utility for the no-institution lottery, her preference over the lotteries will be undefined. Finally, the problem with (4) is that institutions are not necessarily

Given this definition of ex ante efficiency, the claim that ex ante efficiency has moral force emerges as a particular version of the generic claim that rational approvability has moral force.<sup>7</sup> Is the claim true? Imagine that an Agent is choosing among  $\{A_1 \dots A_n\}$ . For some Patient, it would be rational to approve  $A^*$ . Does the Agent therefore have moral reason to perform  $A^*$ ?<sup>8</sup> If so, and if SEU in particular is the correct account of rationality, it follows that ex ante efficiency has moral force. The premise that rational approvability possesses moral weight is intuitively plausible. This premise, or something like it, may well be what motivates the defenders of ex ante efficiency. In any event, the premise bears scrutiny.

## II. DOES RATIONAL APPROVABILITY HAVE MORAL WEIGHT? THE SEU ACCOUNT OF RATIONALITY

In this Part, I consider and reject various arguments why choices by legal officials should, morally, be guided by the rational ordering of those choices by citizens—on the assumption that SEU is the correct account of rationality. Alternative accounts of rationality will be considered below, in Part III. In the analysis that follows, the legal official is the Agent, deciding what policy to put in place, what doctrine to enact, what judgment to issue, etc., while citizens are the Patients with ex ante preferences over the Agent's choices. I use the term "ex ante preference," here and throughout the Article, as shorthand for "maximizes subjective expected utility." The Patient "ex ante prefers"

---

created at canonical times; institutions might come into existence gradually, with no clear answer to the question, "When does the institution start to exist?" Assuming there is a single time at which a given institution comes into existence, (4) is fairly close to my own analysis of ex ante efficiency, and the arguments advanced in this Article could readily be adapted to show that ex ante efficiency as per (4) lacks moral weight.

On these issues, I am indebted to Stephen Perry for extensive, written comments that, among other things, incisively engaged the problem of defining ex ante efficiency and preference. See Stephen Perry, Comments on Manuscript (Mar. 1, 2002) (on file with author).

<sup>7</sup> More precisely, as I explain in the Conclusion, *infra*, the claim that rational approvability has moral force, given SEU, provides one important argument in favor of the moral relevance of ex ante efficiency. This is not the only possible argument. Even if rational approvability lacks moral weight (in general, or on an SEU account of rationality), ex ante efficiency could, in theory, have moral weight.

<sup>8</sup> By "moral reason," of course, I mean a *prima facie* or *pro tanto* moral reason of some kind, not a conclusive moral reason. It would be absurd to suggest that, because some Patient rationally approves  $A^*$ , the morally appropriate choice for the Agent, all things considered, is necessarily  $A^*$ .

some choice by the Agent at  $T$  if and only if it is rational for the Patient, at  $T$ , to approve the choice, given SEU.

### A. *Welfare: Outcomes*

Plausibly, the Agent has moral reason in favor of choices that improve the Patient's welfare. More precisely, the fact that the outcome of some choice  $A_1$  by the Agent is better for the Patient's welfare than the outcome of some other choice  $A_2$  is plausibly a moral consideration for the Agent in favor of  $A_1$ .<sup>9</sup>

Is there, in turn, a link between what is rational for the Patient to approve and her welfare? More precisely, if it is rational for the Patient to approve  $A_1$  over  $A_2$ , and  $A_1$  leads to outcome (possible world)<sup>10</sup>  $W_1$  rather than  $W_2$ , does it follow that  $W_1$  is better for the Patient's welfare than  $W_2$ ? I think not. To begin, there is arguably a disjunction between welfare and preference. Two basic views of welfare are defended within the philosophical literature: preferentialist and substantive. Preferentialists reduce welfare to preference-satisfaction; substantivists, to the attainment of welfare values such as enjoyment, accomplishment, agency, personal relations, and knowledge.<sup>11</sup> On a substantive view of welfare, the Patient might prefer  $W_1$  to  $W_2$  even though  $W_2$  is better for her. For example, her preferences over outcomes might be motivated by altruistic considerations rather than her own welfare. If, further, the Patient has a sufficiently high subjective

---

<sup>9</sup> See, e.g., Matthew D. Adler, *Beyond Efficiency and Procedure: A Welfarist Theory of Regulation*, 28 FLA. ST. U. L. REV. 241, 302-13 (2000) (arguing for moral force of overall well-being).

<sup>10</sup> "Possible worlds" play a substantial role in my analysis. For an introduction to this important concept, see MICHAEL J. LOUX, *METAPHYSICS: A CONTEMPORARY INTRODUCTION* 176-214 (2002). In this Article, I use the term "outcome" to mean a possible world.

<sup>11</sup> On the nature of welfare, see generally JAMES GRIFFIN, *WELL-BEING: ITS MEANING, MEASUREMENT, AND MORAL IMPORTANCE* 7-72 (1986); DEREK PARFIT, *REASONS AND PERSONS* 493-502 (1984); T.M. SCANLON, *WHAT WE OWE TO EACH OTHER* 108-43 (1998); L.W. SUMNER, *WELFARE, HAPPINESS, AND ETHICS* 45-137 (1996); Mozaffar Qizilbash, *The Concept of Well-Being*, 14 ECON. & PHIL. 51 (1998). The hedonic view of welfare might be seen as a type of substantive account—one that makes pleasure the only welfare value. Hybrid views are possible (for example, a view that says an outcome is better if, and only if, both preferred and substantively better) as are other modifications of the basic views (for example, a view that gives welfare weight only to a particular subset of preferences), but for simplicity I consider only the basic views. If, as I argue, the simple preferentialist view doesn't warrant the claim that rational approvability has moral force, then a fortiori hybrid or modified views won't either.



probability that  $A_1$  leads to  $W_1$ , she will rationally choose  $A_1$  and get  $W_1$  notwithstanding the fact that  $W_2$  ranks higher than  $W_1$  with respect to the Patient's own well-being.

Even bracketing the disjunction between preference-satisfaction and welfare, I think it is quite clear that an outcome resulting from the action that the Patient ex ante prefers need not be best for the Patient's welfare. One reason is intuitive. It is a truism that a person can be lucky or unlucky with respect to her welfare. Perhaps moral luck is impossible;<sup>12</sup> but, intuitively, it seems that welfare luck is quite possible indeed. If I pay \$100,000 for a lottery ticket with a prize of \$1 million, where the probability of winning the lottery is 1 in 1 million, and end up winning the lottery, then intuitively I have chosen an action (playing the lottery) that was irrational for me (at least if my choice situation included the option of not playing) but nonetheless resulted in the better outcome for my welfare.

A further reason is given by the structure of SEU itself. SEU drives a wedge between a person's preferences over outcomes and what is rational for a person to choose or approve. Imagine that  $A_1$  will in fact lead to  $W_1$  and  $A_2$  will in fact lead to  $W_2$ , where the person prefers  $W_2$  to  $W_1$ . Within SEU, of course, which action it is rational for the person to choose or approve depends (in part) on her subjective assessment of the likelihood that  $A_1$  or  $A_2$  will result in  $W_1$  or  $W_2$ . If, for example, the person prefers  $W_2$  to  $W_1$ , but (incorrectly) believes it to be highly likely that  $A_1$  will lead to  $W_2$ , and that  $A_2$  will lead to  $W_1$ , then—within SEU—she rationally chooses  $A_1$  even though its outcome,  $W_1$ , is not the outcome she prefers.

One might try to circumvent the objection from SEU itself by distinguishing between "proto-preferences" and "final preferences." Modify SEU as follows: each person has proto-preferences over possible worlds; he rationally chooses or approves actions, depending on his proto-preferences plus his subjective probability assessments; a person's final preferences over outcomes track his rational choices; and a person's welfare ranking of outcomes tracks his final preferences (not his proto-preferences). So, if  $A_1$  leads to  $W_1$  and  $A_2$  leads to  $W_2$ , and our Patient proto-prefers  $W_2$  but rationally approves  $A_1$  because of his limited information and the subjective probabilities grounded therein, the Patient's final preference is for  $W_1$  (the out-

---

<sup>12</sup> For a discussion of the problem of moral luck, with citations to the literature, see Michael S. Moore, *The Independent Moral Significance of Wrongdoing*, 5 J. CONTEMP. LEGAL ISSUES 237 (1994).

come of the ex ante preferable  $A_1$ ) over  $W_2$  (the outcome of the ex ante dispreferable  $A_2$ ). Why not thus salvage the claim that the outcomes of rationally approvable actions are better for those who rationally approve them?

One answer, again, is the intuition that welfare luck is possible. Having welfare track final preferences makes welfare luck impossible.

A more technical point is that a person's final preferences between outcomes are often undefined. A person has a final preference between two possible worlds,  $W_i$  and  $W_j$ , only if the worlds are linked by someone's choice: only if there is some choice situation where one possible action leads to  $W_i$  and another leads to  $W_j$ . For many pairs of possible worlds, there will be no such link.<sup>13</sup> Thus the Patient's final-preference ranking of possible worlds (unlike his proto-preference ranking) will be very "gappy." But it is presumably a desideratum in a theory of welfare that it enable a more, rather than less, complete ranking of outcomes.

Intertemporal variation in preferences poses another obstacle to the putative link between the Patient's final-preference ranking of possible worlds and the goodness of those outcomes for his welfare. Consider  $W_i$  and  $W_2$ , where  $W_i$  would result from choice  $A_1$  by the Agent at time  $T$ ,  $W_2$  would result from choice  $A_2$  by the Agent at time  $T$ , and the Patient ex ante prefers  $A_1$ . The Patient's ex ante preference for  $A_1$  is a product (by definition) of his proto-preference ranking over outcomes (standardly assumed to be temporally fixed), plus his subjective probabilities at  $T$ , linking the choices available to the Agent with the set of possible outcomes. As already mentioned, the Patient's subjective probabilities cannot be assumed to be temporally fixed. They depend on the temporally variable information available to the Patient.<sup>14</sup> But it seems arbitrary to give priority to the Patient's contemporaneous preference with respect to the Agent's choices, over his prior or subsequent preference, in determining the Patient's final-preference ranking of outcomes. Certainly the Patient's contemporaneous preferences have priority in determining what is rational for the

---

<sup>13</sup> Alternatively, for a given pair of possible worlds, could there be multiple choice situations such that, in each situation, one choice leads to one world and another to the other? I'm not sure. If so, a different indeterminacy problem arises, since a given person might rationally choose or approve the choice leading to one world in one situation, but not in the other situation. Just as the existence of *no* choice-link between two worlds creates an indeterminacy in the final-preference ranking, so might the existence of *multiple* choice-links.

<sup>14</sup> On intertemporal variability in subjective probabilities, see, for example, HENRY E. KYBURG, JR., *PROBABILITY AND INDUCTIVE LOGIC* 68-74 (1970).

Patient to choose at a given moment. But we are asking, rather, what it is moral and rational for the *Agent* to choose at a given moment. Insofar as the Patient's preferences should bear on the Agent's choice, it is hard to see why the temporal location of those preferences makes a difference. Informational quality, sincerity, coherence, and other attributes of preference might make a moral difference.<sup>15</sup> However, the correlation between these features and temporal location is contingent, not conceptual, and in the case of informational quality it is subsequent, not contemporaneous, preferences that typically will be best informed.

The point about intertemporal variation will push us away from contemporaneous preferences, and may lead to greater indeterminacy in the Patient's final-preference ranking. At all points in time, the Patient has preferences over the Agent's choices at *T*. The Patient's ex ante (i.e., contemporaneous) preference is for  $A_1$ ; but at other times, the Patient (with different information) may prefer a different action open to the Agent at *T*. One solution is to determine the Patient's final-preference ranking by looking to the Patient's actual preferences at that point in time (not necessarily contemporaneous with *T*) when his preferences most satisfy informational or other criteria. This solution gives moral force to rational approvability by the Patient, but not to ex ante efficiency, since ex ante efficiency as I have conceptualized it is grounded in ex ante preference—in the fact that some choice by the Agent is universally, contemporaneously, and rationally approved by all Patients. Another solution is to withdraw attention from the Patient's actual preferences at any point in time, and look instead to his idealized (say, fully informed) preferences. This moves us away from the SEU account of rational approvability entirely—at least insofar as the idealizing conditions involve full or fuller-than-actual information—since SEU makes it rational for a person to choose or approve the action he prefers given his actual, subjective probabilities, not the action he would prefer under full or fuller information. The third putative solution says that  $W_1$  resulting from  $A_1$  is finally preferred by the Patient over  $W_2$  resulting from  $A_2$ , just in case the Patient at all times actually prefers that the Agent choose  $A_1$ . This is no solution at all, since in many cases intertemporal variation in the Patient's informa-

---

<sup>15</sup> Many philosophers who think that preferences have a role within the correct account of well-being believe that the relevant preferences must be "fully informed" or otherwise idealized. See generally David Sobel, *Full Information Accounts of Well-Being*, 104 ETHICS 784, 790-96 (1994) (discussing accounts of well-being that look to fully informed preferences).

tion will preclude him from having an intertemporally invariant preference for any one choice by the Agent. The upshot of this third "solution" is that the Patient will often lack a final-preference ranking, not just between outcomes unlinked by anyone's choice, but even between outcomes thus linked.<sup>16</sup>

### B. *Welfare: Actions*

The preceding Section relied, implicitly, on a consequentialist view of welfare. I argued that the possible world, or "outcome,"  $W_i$  resulting from an action rationally approvable by some Patient need not be better for the Patient's welfare. I also described problems in generating a "final-preference" ranking of possible worlds that is derivative from the Patient's rational ordering of actions available to an Agent. But why assume that the welfare significance or, more generally, the moral weight of some action is reducible to the goodness of outcomes? Couldn't  $A_1$  be better for the Patient's welfare than  $A_2$ , even though the outcome resulting from  $A_1$  is no better for the Patient than the outcome resulting from  $A_2$ ?

I believe that consequentialism, at least about welfare, is correct, but this is a large and controversial claim that cannot be defended here.<sup>17</sup> More narrowly, consequentialism about welfare and about morality in general is arguably presupposed by SEU—but this narrower claim may be untrue too.<sup>18</sup> Thus, I will entertain the possibility

---

<sup>16</sup> What about a different, consequentialist version of the argument from welfare: not that the rational approvability of  $A^*$  by some Patient entails that  $W^*$  resulting from  $A^*$  is all-things-considered better for the Patient than other outcomes of the Agent's choices, but rather (more weakly) that the rational approvability of  $A^*$  provides the Agent a *prima facie* moral reason, grounded in the Patient's welfare, to choose  $A^*$ ? The claim, in short, would be this:  $W^*$  is in one respect better than the worlds resulting from the alternatives to  $A^*$ , because the Patient rationally approves  $A^*$ . This argument founders on the intrinsic/instrumental distinction, considered below in a different context. *Infra* notes 21-23 and accompanying text. It is true that, if the Patient intrinsically prefers  $A^*$ , and if preferentialism about welfare is true, then  $W^*$  resulting from  $A^*$  is in one respect better than the worlds resulting from the alternative choices. But rational approvability and intrinsic preferability are different things. The Patient might well lack any intrinsic preference for  $A^*$  and approve it purely on instrumental grounds. An action for which a person lacks an intrinsic preference can readily maximize her subjective expected utility, because the possible causal upshots of the action may provide the greatest expected satisfaction of the person's intrinsic preferences, as compared to other actions.

<sup>17</sup> For an excellent collection of philosophical essays debating consequentialism, see *CONSEQUENTIALISM AND ITS CRITICS* (Samuel Scheffler ed., 1988).

<sup>18</sup> This narrower claim seemingly follows from the premise that moral "oughts" generate rational "oughts." Assume that some consideration is a moral reason for the

that both SEU, as an account of rationality, and nonconsequentialism about morality and welfare are true. This brings us back to the suggestion that  $A_1$ 's rational approvability by the Patient might make it better for the Patient's welfare than  $A_2$ —and thereby create a moral reason for the Agent—even though the outcome resulting from  $A_1$  is no better for the Patient than the outcome resulting from  $A_2$ .

In the preceding Section, I contrasted preferentialist and substantive views of welfare. Preferentialist views reduce a Patient's welfare to the satisfaction of the Patient's preferences while substantive views reduce a Patient's welfare to the realization of welfare values such as enjoyment, accomplishment, agency, personal relations, and knowledge. Set aside agency, as there may well be a conceptual link between agency and rational approvability. I think it is quite clear that the remaining entries on the list of plausible welfare values are *not* conceptually linked to rational approvability—at least if rational approvability is specified in line with SEU.<sup>19</sup> Consider enjoyment. The most enjoyable object for the Patient (be it an action or some other object) is not necessarily the object that the Patient most prefers. Enjoyability is a complex hybrid of (1) the subject's positive feelings or physical sensa-

---

Agent to choose  $A_1$  only if the consideration makes it rational for the Agent to choose  $A_1$ , *ceteris paribus*. Now, consider the case of choice by a perfectly informed Agent. Where the Agent is perfectly informed, according to SEU the only considerations that can influence what is rational for the Agent with respect to some choice set of possible actions are considerations that influence his preferences over the possible worlds resulting from those actions. Given some feature of  $A_1$  that supposedly makes it better for the Patient's well-being than  $A_2$ , a perfectly informed Agent choosing in accordance with SEU would not be influenced by this feature of  $A_1$  unless it makes the outcome resulting from  $A_1$  better, in some way, than the outcome resulting from  $A_2$ . However, the premise upon which this argument depends—that moral "oughts" generate rational "oughts"—is itself too large and controversial to defend here.

<sup>19</sup> Different, plausible lists of substantive welfare values are identified in the literature. See, e.g., JOHN FINNIS, *NATURAL LAW AND NATURAL RIGHTS* 90 (1980) (listing seven basic forms of good, including "life, knowledge, play, aesthetic experience, friendship, practical reasonableness, and religion"); JAMES GRIFFIN, *VALUE JUDGEMENT: IMPROVING OUR ETHICAL BELIEFS* 29-30 (1996) (providing a "general profile of prudential values," including "accomplishment," "components of human existence," "understanding," "enjoyment," and "deep personal relations"); GEORGE SHER, *BEYOND NEUTRALITY: PERFECTIONISM AND POLITICS* 199-244 (1997) (defending and expanding upon Derek Parfit's list of goods, that includes "moral goodness, rational activity, the development of one's abilities, having children and being a good parent, knowledge, and the awareness of true beauty" (citation omitted)); Robert E. Goodin & David Parker, *Introduction to Symposium on Martha Nussbaum's Political Philosophy*, 111 *ETHICS* 5, 6-7 (2000) (providing a list of the "central human capabilities," including "life," "bodily health," "bodily integrity," "senses, imagination and thought," "emotions," "practical reason," "affiliation," "other species," "play," and "control over one's environment").

tions; (2) the degree to which the object merits those feelings; and (3) the subject's preference for the feelings. Preference is not the whole story—and if this is true for a relatively subjective value such as enjoyment, it is true *a fortiori* for more objective values such as knowledge, personal relations, or accomplishment. Beliefs must be true (not merely desired) to be instances of knowledge; relationships must be minimally honest, reciprocal, nonabusive, etc., to be valuable; and blade-of-grass-counting or sitcom-watching, however strongly preferred, are not real accomplishments. In short, the substantive welfare values identified by a substantive account of welfare either are realized wholly independent of the welfare subject's preference-satisfaction or at least (with the possible exception of agency) are not reducible to his preference-satisfaction.<sup>20</sup>

Consider, now, the preferentialist view of welfare. Shifting to that view, from the substantive view of welfare, seemingly strengthens the case for the moral (welfare) force of rational approvability. And the combination of (1) preferentialism about welfare; and (2) nonconsequentialism makes that case especially strong. The Patient *ex ante* prefers  $A_1$  to  $A_2$ . With the strictures of consequentialism no longer binding,  $A_1$  might be said to benefit the Patient regardless of how  $W_1$  and  $W_2$  fare in light of her preferences, proto-preferences, or final preferences. With substantive welfare values out of the picture, does it not follow that an action benefits the Patient just in case she *ex ante* prefers it?

That conclusion does not follow. The full-information preferentialist will deny it. And it does not follow even for the preferentialist who looks to actual, rather than fully informed, preferences. The

---

<sup>20</sup> Although there could well be a conceptual link between welfare values and fully informed, or otherwise idealized, preferences, *see, e.g.*, Adler, *supra* note 9, at 298-99 (arguing for such a connection), SEU looks to actual preferences, not idealized ones.

What about the point, to which I allude in the text, that a substantive value might make actual preference-satisfaction a partial condition for the value's realization? Imagine that some action is better for some person's attainment of welfare value  $V$  if, and only if, (1) the person prefers the action; and (2) other conditions obtain? Imagine now that the Agent is faced with a choice of actions, all of which satisfy the non-preferentialist conditions for  $V$  with respect to the Patient. Wouldn't the Patient's preference for  $A^*$  (by triggering the realization of  $V$ ) generate a welfarist reason for the Agent to choose  $A^*$ ? I would respond to this question by adducing my arguments, immediately below, against the moral force of rational approvability given a preferentialist account of welfare. *Infra* text accompanying notes 21-24. If those arguments are sound, then they should also hold good when the welfare account is replaced with one that is substantive but incorporates some values that make preference-satisfaction a partial condition for their realization.

problem of intertemporal variation in the Patient's preferences over the Agent's choices at time *T* looms once more. It is arbitrary to insist that the Patient's *ex ante* (contemporaneous) preference, rather than her subsequent or prior preferences, determine which choice by the Agent at *T* is best for the Patient's welfare. Either we use criteria of welfare-relevance (sincerity, coherence, etc.) to identify the point in time such that the Patient's actual preferences at that point determine which choice by the Agent at *T* is best for the Patient; or we give weight to each moment in the Patient's life and stipulate that the best choice by the Agent at *T* for the Patient's welfare is some function of the set of the Patient's preferences at all moments over the Agent's choices at *T*. In either case, we have abandoned *ex ante* efficiency. And in the latter case, at least, if we refuse to make intertemporal comparisons between the Patient's preferences, there will typically be no determinate answer to the question: "Which choice by the Agent does the Patient prefer?" Such indeterminacy, of course, eliminates rational approvability by the Patient as a possible moral consideration for the Agent.

A different point goes to the distinction between intrinsic and instrumental preference.<sup>21</sup> The Patient might prefer *A*<sub>1</sub> to *A*<sub>2</sub> independently (in part)<sup>22</sup> of their expected causal upshots; or he might prefer *A*<sub>1</sub> to *A*<sub>2</sub> just by virtue of their expected causal upshots. Sweet singing by the Agent, say, might be preferable for the Patient because he intrinsically prefers sweet to sour singing—because sweet singing is something he just likes—or instead because the Patient believes that sour singing is likely to disturb the Patient's sleeping child. In the latter, wholly instrumental case, it is hard to see why the Patient's *ex ante* preference has moral weight for the Agent—even within a preferentialist and nonconsequentialist view of welfare. Imagine that the Agent is an expert in pediatric audiology, who knows that sour, not sweet, singing tends to soothe sleeping children. Should the Agent, seeking to advance the Patient's welfare, sing sweetly or sourly?<sup>23</sup>

---

<sup>21</sup> On the intrinsic/instrumental distinction, see, for example, THOMAS CARSON, *VALUE AND THE GOOD LIFE* 155-58 (2000).

<sup>22</sup> I say "in part" because, even where a person has an intrinsic preference for some action, his decision to choose or approve that action should, rationally, be grounded not merely on that intrinsic preference, but also on his intrinsic preferences for or against the possible causal upshots of the action. If I intrinsically like sweet singing, but believe that this particular instance of sweet singing will trigger a bomb that will injure me, then I rationally should not choose or approve the singing.

<sup>23</sup> Note that, even where the Patient intrinsically prefers some action by the Agent, it remains hard to see why the rational approvability of that action has moral weight for

The case for singing sweetly, I suppose, rests on a robust antipaternalism: (1) preferentialism makes each person sovereign with respect to his welfare; and (2) individual "sovereignty" here means that an individual's welfare is mediated not just by his intrinsic preferences, but by his intrinsic preferences plus his beliefs (correct or not), i.e., promoting someone's welfare means satisfying her intrinsic *or* instrumental preferences. Antipaternalism this robust is not incoherent, but it strikes me as counterintuitive—and much stronger, I should note, than the variants of antipaternalism most widely accepted within the philosophical literature.<sup>24</sup> Antipaternalism this robust is particularly counterintuitive, and perhaps borderline incoherent, in the case where the Patient himself is "deferential": where he prefers that suitably characterized experts give weight to their own subjective probabilities, not the Patient's, in advancing the Patient's welfare. The incoherence could be avoided by weakening the antipaternalist view just described to state that the *ex ante* preferences of nondeferential Patients—whether these *ex ante* preferences are wholly instrumental or, rather, are grounded in part in an intrinsic preference for the action preferred—have moral force for Agents. This is more plausible—but note that the claim that *ex ante* preference, *per se*, possesses moral force has been abandoned.

### C. Agency: Outcomes

What if we focus on the Patient's agency, rather than her welfare? Plausibly, the Agent has moral reason to advance the Patient's agency:

---

the Agent. What has moral weight is the intrinsic preference, not the rational approvability ("*ex ante* preference"), since the latter incorporates probability assessments that the Agent, if more expert in his causal judgments than the Patient, should seemingly discount. See *supra* note 16 (distinguishing between intrinsic preferability and rational approvability). This problem for the putative force of rational approvability emerges particularly clear in the kind of case, discussed in the text, where the Patient has no intrinsic preference for the action rationally approved. However, the problem for the putative force of rational approvability is really more general than that.

<sup>24</sup> For discussions of paternalism, see GERALD DWORKIN, *THE THEORY AND PRACTICE OF AUTONOMY* (1988); JOHN KULTGEN, *AUTONOMY AND INTERVENTION: PARENTALISM IN THE CARING LIFE* (1995); DONALD VANDEVEER, *PATERNALISTIC INTERVENTION* (1986); Joel Feinberg, *Legal Paternalism*, 1 CAN. J. PHIL. 105 (1971). The view is stronger in the following sense: most antipaternalists do not doubt that an Agent can improve a Patient's welfare by performing actions the Patient does not want performed (specifically, by rejecting the Patient's probability assessments and acting in the teeth of the Patient's instrumental preferences). Rather, antipaternalists generally claim that Agents are precluded from performing such actions on liberty, or other nonwelfarist, grounds.



either as one aspect of welfare (on a substantive rather than preferentialist account of welfare which makes agency one of various substantive values), or independent of welfare. Assume, for the moment, that consequentialism about morality is correct. Then the Agent's moral reason to advance the Patient's agency is a moral reason to produce possible worlds that are better in light of the Patient's agency.

What makes one possible world,  $W_i$ , better than another,  $W_j$ , in light of the Patient's agency? Advancing someone's agency could mean advancing her (1) capacity for choice; (2) range of choice; (3) extent of choice; (4) quality of choice; or (5) a hybrid of these, to list some salient possibilities.  $W_i$  is better than  $W_j$  with respect to the Patient's choice-*capacity* if she is better educated, smarter, more thoughtful, and so on in  $W_i$  than in  $W_j$ .  $W_i$  is better than  $W_j$  with respect to the Patient's choice-*range* if the totality of sets of possible worlds accessible from her choice situations in  $W_i$  is better for the Patient's welfare (given a welfare measure applicable to sets of outcomes and appropriate for this context, such as the average or maximum welfare) than the totality of sets of possible worlds accessible from her choice situations in  $W_j$ .  $W_i$  is better than  $W_j$  with respect to the Patient's choice-*extent* if she makes more choices in  $W_i$  than in  $W_j$ .  $W_i$  is better than  $W_j$  with respect to the *quality* of the Patient's choices if the percentage of her choices in  $W_i$  that are rational is higher than the percentage of her choices in  $W_j$ .

Clearly, the link between the Patient's rational ordering of the Agent's actions, and the agency-ranking of the outcomes of those actions (with respect to choice-capacity, -range, -extent, -quality, or some hybrid), is quite contingent. Consider capacity. The Patient might rationally, but unluckily, approve a choice by the Agent that leaves the Patient dead, comatose, impulsive, stupid, or otherwise impaired in her capacity for choice. Similar points hold true regarding range, extent, and quality. First, the Patient's preferences over outcomes might not track their goodness for her agency. Second (and this kind of point should be very familiar by now) even the agency-maximizing Patient might rationally, but unluckily, choose actions with outcomes that diminish agency—given the role of subjective probabilities in determining what is rational within SEU.

The proponent of ex ante preference might try to draw a tighter link between rational approvability and agency. She might say that  $W_i$  is better than  $W_j$  for the Patient's agency just in case, for all choice situations by Agents where  $A_i$  would result in  $W_i$  and  $A_j$  would result in

$W_j$ , the Patient rationally approves  $A_i$  over  $A_j$ . This is, in effect, the final-preference gambit, which I have already discussed and rejected.<sup>25</sup>

#### D. Agency: Actions

In discussing the welfare-relevance of rational approvability, I considered first consequentialist and then nonconsequentialist arguments for a link between the Patient's ex ante preference and his welfare. Agency, too, must be considered in both consequentialist and nonconsequentialist variants. The nonconsequentialist variant says this: where the Patient rationally approves  $A_i$  over  $A_2$ , that fact provides the Agent a reason (grounded in the Patient's agency) to choose  $A_i$ , regardless of whether the outcomes resulting from  $A_i$  would be better for the Patient's agency.

This argument will perhaps seem more familiar and persuasive if phrased in the language of consent. The Patient rationally approves  $A_i$  over  $A_2$ . In effect, he consents to  $A_i$  over  $A_2$ . But consent has moral force. Therefore the Agent has moral reason (nonconsequentialist in structure) to choose  $A_i$  over  $A_2$ .

Does the argument work? Note that it rests on the Patient's hypothetical, not actual, consent. The Agent is considering  $\{A_1 \dots A_n\}$ .  $A^*$  is rational for the Patient to approve; specifically, since SEU is our current theory of rationality,  $A^*$  is the possible action that the Patient ex ante prefers. This feature of  $A^*$  does not entail that the Patient has actually chosen  $A^*$ . A person can ex ante prefer an option without actually choosing that option, indeed without thinking about the option at all. The Patient's ex ante preference for  $A^*$  is inferred from his *preferences* over possible worlds, plus his *subjective probabilities* over possible worlds conditional on  $A^*$ . Both preferences and subjective probabilities (as beliefs) are dispositional. They entail what the Patient *would choose*, but exist at all times—and generate ex ante preferences over contemporaneous choices by other Agents at all times—whether or not the Patient actually engages in choice and thought.<sup>26</sup>

Note further that a Patient who is actually engaged in choice contemporaneous with the Agent's choice—who is consciously engaged in deciding whether to approve  $\{A_1 \dots A_n\}$ —need not pick  $A^*$ , even

<sup>25</sup> *Supra* text accompanying notes 13-16.

<sup>26</sup> See DANIEL M. HAUSMAN & MICHAEL S. MCPHERSON, ECONOMIC ANALYSIS AND MORAL PHILOSOPHY 28 (1996) ("We regard  $Q$ 's preference ranking as a subjective state of  $Q$  that, along with  $Q$ 's beliefs, explains her choices."); JAEGWON KIM, PHILOSOPHY OF MIND 158-59 (1996) (noting that beliefs and desires need not be occurrent).

though  $A^*$  is the rational choice for the Patient to approve. After all, the Patient's actual choices can be irrational. A person's rational ordering is at best a fallible predictor of his actual choices.

Rational approvability is logically linked to *hypothetical*, rather than actual, consent—specifically, to hypothetical rational consent. To say that the Patient ex ante prefers  $A^*$  is to say that the Patient would choose  $A^*$ , if hypothetically he were faced with the choice of approving  $A^*$  or the other options and were to choose rationally in accordance with SEU. But why think that this kind of hypothetical consent has moral force? Some theorists insist that no hypothetical consent has moral force, rather that consent must be actual to be morally powerful.<sup>27</sup> It is also plausible to think that hypothetical, fully informed consent, in addition to (or in lieu of) actual consent, has moral force.<sup>28</sup> Crucially, however, the Patient's hypothetical consent to  $A^*$  is *not* fully informed. The Patient would hypothetically approve  $A^*$  were she to choose rationally, given her subjective probabilities. But those are in turn grounded in whatever incomplete stock of information the Patient possesses at the time of the Agent's choice. A person's subjective probabilities are numerical representations of her beliefs (where 1 represents a belief she holds with certainty, 0 represents no belief, and a number between 0 and 1 represents a belief she holds with some degree of uncertainty). There is absolutely no requirement in SEU, or in the concepts of ex ante preference and ex ante efficiency, that the person holding such probabilistic beliefs be fully informed. Were there such a requirement, the distinction between ex ante and ex post preference—between preferences over actions and preferences over outcomes—would dissolve, at least in a universe operating under deterministic laws. Fully informed persons would have a subjective probability of 1 that each  $A_i$  would result in its actual upshot,  $W_i$ , and a subjective probability of 0 that each action would result in any other outcome. But, of course, subjective probabilities within SEU can deviate from both 0 and 1. This deviation (in a deterministic universe) reflects the deviation between the ex ante preferences of a fully informed Patient—whose probabilities would be

---

<sup>27</sup> See, e.g., Heidi M. Hurd, *Justifiably Punishing the Justified*, 90 MICH. L. REV. 2203, 2305 (1992) ("[H]ypothetical consent lacks just the element that makes an act of consent morally significant; namely, consent.").

<sup>28</sup> Cf. MICHAEL SMITH, *THE MORAL PROBLEM* 151-52, 155-61 (1994) (analyzing normative reasons in terms of "fully rational" desires, where "full rationality" includes full information).

only 1s and 0s—and the *ex ante* preferences of an actual, possibly partially uninformed, Patient.<sup>29</sup>

Nor can the supposition that the universe is indeterministic salvage the argument from the moral force of hypothetical fully informed consent to the moral force of *ex ante* preference. Fully informed individuals would have subjective probabilities that track the objective, physical probabilities.<sup>30</sup> There is no requirement in SEU that individuals' probabilities track the objective, physical probabilities. To insist on that is to abandon SEU and erect instead some variant of objective expected utility—a theory of rationality where the rational choice for an actor depends on relative frequencies or on the objectively warranted partial beliefs (objective probabilities), rather than her actual partial beliefs (subjective probabilities). SEU captures the thought that what is rational for a person—as opposed to what is most valuable for her, or what is the right thing for her to do—depends on her actual beliefs and preferences, which can deviate from objective truths or values. Neither in a deterministic nor an indeterministic universe does SEU stipulate that a given Patient's probability assignment must match the probability assignment of a fully informed Patient.

Hypothetical consent might be morally powerful even if uninformed. Rawls's veil of ignorance provides a famous example. For Rawls, hypothetical choice under conditions of radical informational limitations is the mark of justice.<sup>31</sup> But Rawls's choosers are systematically uninformed—stripped of all information about their personal circumstances and endowments—so that their rational choices track what would be morally reasonable for them to choose. The Patient in our scenario is arbitrarily uninformed. He has whatever information he has managed to accumulate up until the point in time when the Agent makes her choice. Based on that historically contingent stock of information—one that has not been structured and restricted in light of some moral theory, such as Rawls's—we determine the Patient's subjective probabilities and, thereby, what he *ex ante* prefers. Yet it is very hard to see why hypothetical consent by arbitrarily unin-

---

<sup>29</sup> Cf. Stephen R. Perry, *Risk, Harm and Responsibility*, in *PHILOSOPHICAL FOUNDATIONS OF TORT LAW* 321, 330-39 (David G. Owen ed., 1995) (discussing connection between probability and determinism).

<sup>30</sup> See *infra* text accompanying note 38 (discussing objective, physical probabilities).

<sup>31</sup> JOHN RAWLS, *A THEORY OF JUSTICE* (1971).

formed persons—neither fully informed, nor uninformed in a structured and morally appropriate way—would possess moral force.

Again, abandoning SEU could be a solution—one we will consider below. Arguably, if all Patients under conditions of morally structured ignorance (such as Rawlsian radical ignorance) would approve some choice by the Agent, then the Agent has moral reason in favor of that choice. Ex ante efficiency might have moral force if defined in terms of ex ante preferences under morally special informational conditions. Notably, however, most scholarly proponents of ex ante efficiency do not assume such conditions (for example, the actor's ignorance of his own cost structure, abilities, or tastes).<sup>32</sup> Relatedly, it is a delicate matter to build morally appropriate informational conditions into a theory of *rationality*. What I am asking, in this Article, is whether the rational approvability for some person of some actor's option creates a moral reason for the actor in favor of the option—not whether the person's hypothetical consent to the option under certain conditions of partial information does. In any event, given the standard SEU account of rationality, rational approvability lacks moral force qua hypothetical consent because there need be no special structure to the Patient's beliefs whatsoever (beyond their conforming to the probability calculus).

### E. *Stability*

Sometimes a link is drawn between ex ante efficiency and the value of stability.<sup>33</sup> Assume the Agent at time *T* is choosing between *A*<sub>1</sub> and *A*<sub>2</sub>, where *A*<sub>1</sub> is universally ex ante preferred. Then *A*<sub>2</sub> is, supposedly, unstable. Were the Agent to choose *A*<sub>2</sub>, everyone would agree to override that choice and choose *A*<sub>1</sub> instead. Or so the argument goes.

Note that other actors could not literally reverse the Agent's choice of *A*<sub>2</sub>. *A*<sub>2</sub> is a particular action (bodily movement) by the Agent which, once performed, is an irrevocable feature of the past. Nor

---

<sup>32</sup> For example, none of the authors cited in footnote one, as I read them, presuppose that the partial information generating ex ante preferences, and thereby ex ante efficiency, is itself structured by moral considerations. Farber does assume structured uncertainty à la Rawls, but only at the threshold stage of choice between legal principles—which, he argues, eventuates in the principle of ex ante efficiency—and not at the later points where this principle is applied by judges.

<sup>33</sup> Cf. Matthew D. Adler & Eric A. Posner, *Implementing CBA When Preferences Are Distorted*, in *COST-BENEFIT ANALYSIS: LEGAL, ECONOMIC AND PHILOSOPHICAL PERSPECTIVES* 269, 300-05 (Matthew D. Adler & Eric A. Posner eds., 2001) (discussing problem of "market adjustment" to governmental approval of inefficient programs).

could other actors literally restore the possible world  $W_i$  that would have resulted had the Agent chosen  $A_i$  at  $T$ .  $W_i$  involves the Agent doing  $A_i$  at  $T$ ; by hypothesis, the Agent has not done that.

Rather, the claim must be that there is some collective action  $A_c$  which the Patients could choose at some time after  $T$  and which is practically indistinguishable from  $A_i$ . The Patients rationally would choose  $A_c$  if the Agent chooses  $A_2$ —thus, in effect, “overriding”  $A_2$ . However, there may be no such collective action available to the Patients and rational for them to choose: both because of the potential physical effects of  $A_2$  (if the Agent drops a nuclear bomb at  $T$ , what can the Patients do to reverse that?), and because of the potential change in the Patients’ information after  $T$ . Further, even if some such collective action,  $A_c$ , exists, collective action problems or irrationality on the part of some or all of the Patients may prevent them from agreeing on  $A_c$ .

The moral Agent, choosing between  $A_1$  and  $A_2$ , will consider the possible upshots of these actions—including the possibility that  $A_2$ , if universally dispreferred, will result in some subsequent collective choice  $A_c$  that the Agent views as practically indistinguishable from, or worse than, the original  $A_i$ . But the fact that  $A_i$  is universally ex ante preferred to  $A_2$  does *not* imply that some such  $A_c$  would occur, or that the Agent would predict it to occur. Ex ante inefficient choices are not, necessarily, actually or predictably unstable.

#### F. *Official Uncertainty*

Up to this point, I have assumed that the legal official—our Agent—is perfectly informed, although the citizens whose preferences and interests she considers—the Patients—are not. What I have shown, in effect, is that the decision in any choice situation by an Agent who is (1) solely motivated by moral considerations; and (2) perfectly informed would not be influenced by the rational ordering of the Agent’s choices by any Patient (if SEU is the correct account of rationality). But what if we relax the assumption of perfect information on the Agent’s part? Isn’t the Patient’s rational ordering an *indicator*, for an uncertain Agent, of what the Agent ought to do, in light of one moral consideration—the Patient’s welfare?

This raises the complex issue of moral decision making under uncertainty. For simplicity, and to make a strong case for the moral force of ex ante preference, I will assume the following: morality is consequentialist in structure, and the morally appropriate action un-

der uncertainty is to maximize subjective or objective *expected moral goodness*. That is: (1) each choice,  $A_i$ , by the actor can be represented as a lottery over possible worlds (where the probabilities in the lottery are construed as representing either the actor's beliefs or the objective probabilities, depending on whether the expected moral goodness account is subjective or objective); (2) the comparative moral goodness of possible worlds can be represented by a numerical "goodness" function; (3) for each choice, an *expected moral goodness* number can be calculated by using the probabilities in the matching lottery to discount the goodness of the various possible outcomes and then aggregating; and (4) the morally appropriate choice is the choice with the largest expected goodness.<sup>34</sup>

Why does this account of moral choice under uncertainty generate a strong case for the moral force of rational approvability? It does so by creating an isomorphism between (1) the moral ordering, by an uncertain Agent, of the choices available to him; and (2) the rational ordering, by an uncertain Patient, of the choices available to the Agent. The Agent orders the choices in light of their expected moral goodness; the Patient, in light of their subjective expected utility. Expected utility and expected moral goodness are methods for ranking choices that start with a cardinal function of possible worlds (a utility or goodness function), and then assign each choice a number equaling the probabilistically weighted average of the cardinal numbers assigned to all of the choice's possible outcomes.

Clearly, the Agent's expected-moral-goodness ranking of his choices, and a given Patient's subjective-expected-utility ranking of the same set of choices, can diverge once moral considerations other than the Patient's welfare come into play. But the following Convergence Claim is tempting: an uncertain Agent's expected-moral-goodness ranking of his choices, *in light of the Patient's welfare*, and the Patient's subjective-expected-utility ranking of those choices, are the same ranking. Assume the Convergence Claim is true. Imagine, further, that a given choice,  $A^*$ , by an uncertain Agent is ex ante efficient;  $A^*$  maximizes expected utility for all Patients. Imagine, finally, that welfare is the only relevant moral consideration. Then it would follow that  $A^*$  is the morally appropriate (expected-moral-goodness maximizing) choice for this uncertain Agent. In short, the Convergence Claim im-

---

<sup>34</sup> See FUMERTON, *supra* note 2, at 90-128 (discussing "value adjusted possible consequence act consequentialist" conceptions of both rational and morally right action).

plies that uncertain Agents morally should choose ex ante efficient actions, at least within a morality that is consequentialist and welfarist.

So is the Convergence Claim true? There are two important considerations that undermine the claim, one arguable and one quite certain. First, the substantive rather than preferentialist account of welfare is arguably correct. On the preferentialist account, the Agent's ranking of possible worlds in light of the Patient's welfare will necessarily track the Patient's own preferences over those outcomes. On the substantive account, however, it need not. Second, and more crucially, even if the preferentialist account of welfare is correct—so that the Agent, focusing on the Patient's welfare, can equate his moral goodness function with the Patient's utility function—it does not follow that the expected moral goodness for a given choice equals its subjective expected utility. For the Agent, as for the Patient, each choice,  $A_i$ , by the Agent is a lottery across possible worlds. *But the probability numbers in the lottery need not be the same.* The numbers used to calculate subjective expected utility are the Patient's subjective probabilities; the numbers used to calculate expected moral goodness are either the Agent's subjective probabilities or objective probabilities, both of which can deviate from the Patient's subjective probabilities.

Here is a silly example that makes the point quite obvious. There is a lake outside the Patient's house. If the lake stays in its natural state, the Patient's utility is 0. If the lake turns green, the Patient's utility is 10. If the lake turns red, the Patient's utility is -10. The Agent has a choice of dumping or not dumping some harmless colorant into the lake. The objective probability that the colorant would turn the lake green is  $1/5$  and that it would turn the lake red is  $4/5$ . The Agent's subjective probabilities are  $3/10$  and  $7/10$ , respectively. However, the Patient's subjective probability of green conditional on dumping is  $3/5$  and of red conditional on dumping is  $2/5$ . The objective expected moral goodness of dumping is -6; the subjective expected moral goodness of dumping is -4; and the subjective expected utility of dumping is 2. The choice by the Agent that is rationally approvable by the Patient is to dump the chemical; but the morally appropriate choice for the Agent in light of the Patient's welfare, whether objective probabilities or the Agent's own subjective probabilities are used to generate the moral ordering, is to refrain from dumping.

The morally appropriate action for an uncertain Agent is surely sensitive to the probability that the various choices open to her will re-



sult in morally better or worse outcomes. But it is an unwarranted leap from this premise to the conclusion that the uncertain Agent ought morally to be influenced by the ex ante preferences of the persons affected by her choices. Plausible normative accounts of choice under uncertainty look either to the actor's subjective probabilities (as does SEU), or to the objective probabilities. But the probability numbers used to determine each Patient's ex ante preference are *different* from both of these; they are, by definition, the Patient's own subjective probabilities. Nor is it plausible to say that the Agent, acting under uncertainty, should rely on the subjective probabilities of the persons whom she affects, since she typically affects multiple persons with potentially quite different subjective probabilities. The gambit of introducing uncertainty on the part of the Agent brings probability into play—but not the probability numbers that determine, within SEU, what is rationally approvable by each Patient.

G. *Conclusion: The Moral Weight of Rational Approvability Within SEU*

SEU ties rationality to subjective probabilities and preferences. Both of these features undermine the claim that rational approvability, as given by SEU, has moral force. Tying rationality to preference-satisfaction breaks the link between rational approvability and welfare—at least if welfare is itself a matter of enjoyment, friendship, accomplishment, knowledge, and other such values, and not of preference-satisfaction. SEU's reliance on subjective probabilities is an even deeper problem. It vitiates the link between the Patient's preferences over outcomes and her preferences over actions, since the Patient's beliefs about which outcomes would result from which actions can be mistaken. It creates intertemporal change in the Patient's preferences over a particular set of actions since what the Patient believes can change over time. Giving special moral weight to the preferences that the Patient happens to hold at a particular time, merely in virtue of their temporal location (e.g., to preferences contemporaneous with the Agent's choices), seems arbitrary. SEU's reliance on subjective probabilities drives a wedge between rational approvability and the varieties of hypothetical consent that arguably have moral force, such as hypothetical, fully informed consent or hypothetical consent under conditions of structured ignorance. Finally, it drives a wedge between what is morally appropriate for an uncertain Agent, and what is rationally approvable by an uncertain Patient. If the Agent maximized expected moral goodness, and the Patient maximized expected utility, using the same *objective* probabilities, the two orderings by these two

uncertain persons would converge (with some further assumptions). SEU precludes this convergence.

Let us turn, then, to consider the moral force of rational approvability within accounts of rationality that drop one, or both, of these problematic features of SEU.

### III. DOES RATIONAL APPROVABILITY HAVE MORAL WEIGHT? ALTERNATIVE ACCOUNTS OF RATIONALITY

#### A. *Objective Expected Utility*

The objective expected utility (OEU) account of rationality amends SEU by replacing subjective with objective probabilities.<sup>35</sup> OEU leaves the other features of an expected utility account in place: the construal of each action as a lottery over possible outcomes; the existence of a utility function representing the actor's preferences over outcomes; and the identification of the rational choice as the choice the expected utility of which (calculated within OEU using objective rather than subjective probabilities) is greatest. I take it that "objective" probabilities are interpersonally—if not intertemporally—objective; that is, where the Agent is choosing among  $\{A_1 \dots A_n\}$  at time  $T$ , the objective probability that some choice,  $A_i$ , will lead to some outcome,  $W_j$ , is the same at  $T$  for all Patients affected by the Agent's choice and for the Agent herself. An expected utility account where probabilities are objective but relativized to particular persons is sufficiently close to SEU that it does not merit separate consideration here. Furthermore, the proponents of OEU, and more generally of objective probability, generally consider probabilities to be interpersonally objective.

Two standard accounts of probability contend within the modern philosophical literature: Bayesian accounts, which construe probability as a degree of belief, a numerical measure of someone's partial belief; and frequentist accounts, which construe probability as the frequency with which objects (including persons, actions, or other objects) in some reference class have some attribute.<sup>36</sup> A Bayesian ac-

---

<sup>35</sup> See EELLS, *supra* note 2, at 31 (suggesting OEU as an alternative account of rationality to SEU).

<sup>36</sup> Good overviews of the philosophical literature on probability include L. JONATHAN COHEN, *AN INTRODUCTION TO THE PHILOSOPHY OF INDUCTION AND PROBABILITY* (1989); DONALD GILLIES, *PHILOSOPHICAL THEORIES OF PROBABILITY* (2000); HENRY E. KYBURG, *PROBABILITY AND INDUCTIVE LOGIC* (1970); STORRS

count that looks to the actual partial beliefs that persons possess generates subjective, not objective, probabilities: the subjective probabilities that figure in SEU. A Bayesian account that looks to the partial beliefs that persons ought to possess, given their evidence, generates probabilities that are objective but not interpersonally—or intertemporally—so.<sup>37</sup> The Bayesian might, however, generate interpersonally—and intertemporally—objective probabilities by looking to the partial beliefs that an idealized person, in possession of full information or a defined stock of information, would have. Where the Agent at  $T$  is choosing among  $\{A_1, \dots, A_n\}$ , the objective probability that choice  $A_i$  leads to outcome  $W_j$  is the numerical degree of belief that an idealized person, with full or defined information, would attribute (at all times) to the proposition " $A_i$  results in  $W_j$ ."

Turning to frequentism, a crucial issue is how to specify the reference class within which to subsume any given action, so as to enable a calculation of the frequency with which the action results in some type of outcome. Consider the Agent's action,  $A_1$ , of placing benzene in a lake near the forty-year-old Patient's home. What is the frequentist probability that this action will result in the Patient's death from cancer?  $A_1$  might be variously classified as "exposing a person to a toxin," "exposing a person to benzene," "exposing a person to benzene that is dissolved in water," "exposing a forty-year-old person to benzene," and so on. The frequency with which actions in the first class result in cancer deaths differs from the frequency with which actions in the second class do, and so forth. Frequentists differ in their proposals as to the correct reference class for determining the probability that a particular object (here, the particular action of placing benzene in the lake) has an attribute (here, the attribute of being carcinogenic). Some propose that the correct reference class is the *conventionally salient* class (the class in which ordinary members of the relevant society

---

MCCALL, A MODEL OF THE UNIVERSE: SPACE-TIME, PROBABILITY, AND DECISION 141-47 (1994); ROY WEATHERFORD, PHILOSOPHICAL FOUNDATIONS OF PROBABILITY THEORY (1982); Colin Howson, *Theories of Probability*, 46 BRIT. J. PHIL. SCI. 1 (1995). For presentations of the "probability calculus"—the numerical axioms that any account of probability, be it Bayesian, frequentist, or other, must satisfy—see IAN HACKING, AN INTRODUCTION TO PROBABILITY AND INDUCTIVE LOGIC 23-78 (2001); RESNIK, *supra* note 2, at 45-80; BRIAN SKYRMS, CHOICE AND CHANCE: AN INTRODUCTION TO INDUCTIVE LOGIC 130-59 (2d ed. 1975).

<sup>37</sup> Of course the terms "subjective" and "objective" are slippery. However, if the beliefs of a person idealized in one way or another are seen to be suitably "objective"—in that they can differ from the actual beliefs that person happens to possess—then Bayesian (i.e., degree-of-belief) accounts of probability can generate objective probabilities.

would subsume the object); others that it is the *statistical* class (the narrowest class for which reliable frequency information is available); and still others that it is the *physical* class (the class consisting of objects sharing all the causally relevant characteristics of the object in question).<sup>38</sup> Using conventionally salient or statistical reference classes will produce frequentist probabilities that are interpersonally, but not intertemporally, objective while using physical reference classes will produce frequentist probabilities that are both interpersonally and intertemporally objective.

If rational choice and approval is defined by OEU rather than SEU, does the fact that some action,  $A^*$ , by an Agent at time  $T$  is rationally approvable by a Patient at  $T$  create a moral reason for the Agent to choose that action? A crucial obstacle here is that OEU incorporates a numerical ranking of outcomes, "utility," that depends on the Patient's *preferences* over the outcomes. If a substantive rather than preferentialist account of welfare is correct—and there are plausible reasons to believe this is the case<sup>39</sup>—the fact that the Patient prefers outcome  $W_1$  to  $W_2$  neither (1) means that  $W_1$  is better than  $W_2$  for the Patient's welfare; nor (2) means that  $W_1$  is better for the Patient in some other way (for example, by enhancing his agency). Even if objective probabilities reflect full information,<sup>40</sup> and a fortiori if objective probabilities are defined in some other way,  $A^*$  might be rationally approvable by the Patient but not lead to the outcome best for her welfare or best for her in some other way. Nor, on a substantive account of welfare, will it necessarily be true that the morally appropriate choice for the uncertain Agent, in light of the Patient's welfare, is  $A^*$ —however objective probabilities are defined. The morally appropriate choice for the uncertain Agent, in light of the Patient's welfare, is a product of a goodness function, that ranks outcomes with respect to the Patient's welfare, plus the probabilities (in some sense) of the various actions available to the Agent leading to the various possible outcomes. But on a substantive account of welfare, this goodness function need *not* track the Patient's utility function.

---

<sup>38</sup> For a discussion of the role of reference classes in determining frequentist probabilities, and of these particular types of classes, see Matthew D. Adler, *Risk, Death and Harm: The Normative Foundations of Risk Regulation*, 87 MINN. L. REV. (forthcoming 2003).

<sup>39</sup> See, e.g., GRIFFIN, *supra* note 19, at 21-29 (advancing considerations against the "taste model" of well-being).

<sup>40</sup> Objective probabilities would reflect full information if they were defined in Bayesian fashion as the degrees of belief of fully informed persons, or in frequentist fashion as frequencies relative to physical reference classes.

Shift, now, to a preferentialist view of welfare. The fact that  $A^*$  is rationally approvable by the Patient still does not entail that  $W^*$ , which in fact results from  $A^*$ , is best for the Patient. Welfare luck is possible. The only way to make the rational approvability of  $A^*$  within OEU imply the welfare goodness of  $W^*$  is (1) to have the objective probabilities reflect full information; and even then only if (2) the universe is deterministic (so that fully informed persons would always know for certain which outcomes will result from which actions). But full-information objective probabilities, in a deterministic universe, will always be 0 or 1, never in between. Thus the version of OEU just described violates two truisms of welfare and rationality: it makes welfare luck impossible (i.e., the rationally approvable action for the Patient necessarily produces the outcome that is best for her welfare), and it precludes probabilities greater than 0 and less than 1 (but surely the rational choice is sometimes a choice that reflects objective or subjective *uncertainty*).<sup>41</sup>

Let us focus, now, on the choices of an Agent who is himself uncertain. If an uncertain Agent's morally appropriate course is to maximize expected moral goodness—using the *same* interpersonally objective probabilities that determine the Patient's expected utility—and if the Agent's moral goodness ranking of outcomes, in light of the Patient's welfare, tracks the Patient's utility ranking of outcomes, then it follows that: the choice  $A^*$  that maximizes the Patient's objective expected utility is necessarily also the choice that maximizes objective expected moral goodness, in light of the Patient's welfare. This is true regardless of how objective probabilities are defined—even if, most plausibly, they do not reflect full information. In short, the morally appropriate choice for the uncertain Agent, in light of the Patient's

---

<sup>41</sup> In addition, the "final-preference" gambit is not strengthened by the shift to OEU. My objections to that gambit within SEU were: (1) it makes welfare luck impossible; (2) a person's final-preference ranking of pairs of worlds will often be undefined, specifically where the worlds are not linked by some actor's choice; and (3) a person's probability assessments vary over time, and giving weight to her probability assessments contemporaneous with the choice situation generating the final-preference ranking seems arbitrary. The first and second problems are not avoided by any version of OEU; the third is only avoided by a version that precludes intertemporal variation in objective probabilities.

The nonconsequentialist variant of the welfare argument for the moral force of rational approvability, *supra* Part II.B, is also not substantially bolstered by the shift from SEU to OEU. At least for what would seem to be the most plausible specifications of objective probability, namely frequentist probabilities relative to conventional or statistical reference classes, the problems facing SEU—intertemporal variation in preference and the questionable moral relevance of the instrumental calculations underlying the Patient's ex ante preference for some action by the Agent—also face OEU.

welfare, and the choice rationally approvable by the Patient will be the same. The move from SEU to OEU permits this convergence by eliminating the possible deviation between the probabilities relevant for the Agent's moral choice and the probabilities relevant for the Patient's rational approval. But the convergence, crucially, will not occur except on a preferentialist view of welfare—a view that, as I have said, is problematic.

Moving away from considerations of welfare, and from a consequentialist moral framework, the proponent of the moral force of rational approvability might point to hypothetical consent. If objective probabilities reflect full information, then the fact that the Patient rationally approves *A\** implies that the Patient would consent to *A\** under full information. If objective probabilities are defined, alternatively, as the degrees of belief of persons possessing a morally ideal stock of partial information—for example, persons in a state of radical Rawlsian ignorance—then the fact that the Patient rationally approves *A\** implies that the Patient would consent to *A\** if partially informed in this ideal way. But there are real difficulties in equating the probabilities that bear on *rational* choice with the probabilities of fully informed, or ideally uninformed, persons. Here is yet another plausible truism of rationality: what is rational for a person depends upon her actual beliefs, or at least on her actual evidence (for example, on the beliefs she is warranted in holding given that evidence), or at least on the evidence she would possess had she gathered evidence appropriately. As Richard Fumerton, in criticizing the view that equates rational choice with choice under full information, observes:

If I were to know all of the information that bears on some decision, I might well act in ways that in fact I have every reason to believe would be disastrous. But surely we want to make the rational . . . course of action for me to take a function of my epistemic situation.<sup>42</sup>

The definitions of objective probability just considered, which create a link between rational approvability and hypothetical consent under morally significant conditions, radically depart from the actor's own "epistemic situation." More generally, Fumerton's point calls into question all OEU accounts—although there could be such accounts where the deviation from the actor's epistemic situation is less pronounced, for example, where objective probabilities are defined in frequentist fashion and relative to conventional or statistical rather than physical classes.

---

<sup>42</sup> FUMERTON, *supra* note 2, at 102-03.

In sum, although the issues are complex, it does not seem that there is an OEU account of rationality that both (1) is in fact a plausible account of what rational choice consists in, and (2) makes the rational approvability, by some Patient, of some Agent's choice, a morally relevant consideration for that Agent.

#### B. *Subjective Expected Value and Objective Expected Value*

The subjective expected value (SEV) and objective expected value (OEV) accounts of rationality suppose that there is a cardinal index of the value of different outcomes,  $(V(W_i))$ , which plays a role in determining the rational choice parallel to the role that the utility index plays within SEU.<sup>43</sup> Each possible action can be represented as a lottery across possible worlds  $\{p_1 \dots p_n\}$ , with the probabilities given by a measure of the actor's actual beliefs (SEV), or by some objective measure (OEV). SEV and OEV then specify the rational action as that action the expected value of which— $p_1 * V(W_1) + p_2 * V(W_2) + \dots + p_n * V(W_n)$ —is greatest. For simplicity, assume that the value index measures the *value for the actor* (the welfare-value, agency-value, or some such sense) of the various possible outcomes.

On an SEV account of rationality, does the rational approvability of some Agent's choice by a Patient create a moral reason for the Agent? I suggest not. I argued in Part II that rational approvability, within SEU, lacks moral weight—both because the numerical index (utility) of outcomes within SEU does not track their welfare-value for the Patient, and on various other grounds. Shifting from SEU to SEV potentially eliminates the divergence between the numerical index of outcomes (now a value index) and their welfare-value for the Patient; it does not, however, eliminate the other objections to the claim that rational approvability has moral weight. Note the following objections that still obtain: (1) because he does not know for certain which outcomes result from which actions, a Patient might rationally approve an action the outcome of which is worse for him than the outcome of some alternative action open to the Agent; (2) the gambits of responding to this objection by creating "final preferences" over outcomes that necessarily track the Patient's rational approvals, or by shifting to a nonconsequentialist construal of welfare, are unsuccessful.

---

<sup>43</sup> See EELLS, *supra* note 2, at 31 (presenting OEV and SEV accounts of rationality as alternatives to SEU).

ful;<sup>44</sup> (3) appeal to considerations of agency does not work, again because the Patient lacks certain knowledge, and might maximize subjective expected value by choosing an action that results in an outcome which is worse for his agency than alternatives; (4) hypothetical consent does not do the trick, since the Patient's subjective probabilities are based on his actual stock of information, rather than reflecting full information or a morally ideal stock of partial information; and (5) the choice that maximizes SEV for the Patient is not necessarily the morally appropriate choice for an uncertain Agent, since that choice would depend either on the Agent's own subjective probabilities (which can deviate from the Patient's) or on the objective probabilities.

OEV circumvents some, not all, of these objections. It bolsters an argument from hypothetical consent *if* objective probabilities are defined as the probabilities that persons possessing full information, or a morally ideal stock of partial information, would possess. But as I argued in the prior Section, there are real difficulties in equating the probabilities that bear on *rational* choice with the probabilities of fully informed, or ideally uninformed, persons. OEV also serves to tighten the link between what is rationally approvable for a Patient and what is morally appropriate for an uncertain Agent in light of the Patient's welfare. *However* objective probabilities are defined, within OEV, if the appropriate choice for the uncertain Agent is to maximize expected moral goodness *using these objective probabilities*, then rational approvability *does* have a kind of moral weight for the Agent. Note, crucially, that the argument does not depend on specifying objective probabilities as reflecting full information, or morally ideal partial in-

---

<sup>44</sup> The "final-preference" gambit within SEV would say that a Patient has a "proto-value" ranking of worlds, measured by a numerical value index, and that the world actually resulting from a choice maximizing subjective expected proto-value for the Patient is better for her than the world resulting from another choice in the same choice situation. But this gambit (as within SEU) makes welfare luck impossible; it means that the final welfare ranking of two worlds, for a person, might often be undefined; and it runs up against the problem of intertemporal variation in subjective probabilities.

The nonconsequentialist welfare gambit within SEV would say that the SEV-maximizing action by the Agent is better for the Patient's welfare than alternative actions, even if the world resulting from that action is not better than the worlds resulting from the alternative actions. Here, too, the objections considered above, in analyzing SEU, are not dissipated. The SEV-maximizing action by the Agent as determined using the Patient's subjective probabilities at the time of choice might not be the SEV-maximizing action as determined using the Patient's subjective probabilities at some other time. And an action might maximize SEV for the Patient but simply have instrumental, not intrinsic, welfare value; if so, the (fully informed) Agent would seemingly have no moral reason to take that action.



formation. Imagine that objective probabilities are, instead, frequencies relative to statistical reference classes; if these probabilities both drive rational choice and moral choice under uncertainty, then the argument for the moral weight of rational approvability goes through.

But is it plausible that there are interpersonally objective probabilities which determine rational choice (hence are not defined in terms of full information or morally ideal partial information) and *also* determine moral choice? The problem of moral choice under uncertainty has not been discussed at great length in the philosophical literature,<sup>45</sup> and it is beyond the scope of this Article to explore that problem in detail, much less propose a particular account. Let me merely note some possible accounts—for simplicity, ones that could be proposed within a consequentialist moral framework. One conceivable view is that the morally appropriate choice, regardless of the Agent's information, is the choice that produces the best outcome. Another is that the morally appropriate choice is the choice that a fully informed Agent, maximizing expected moral goodness, would choose.<sup>46</sup> Yet another view is that the morally appropriate choice is the choice that the Agent, with her actual subjective probabilities, maximizing subjective expected goodness, would choose. Yet another is this: we calculate the probabilities that the Agent ought to ascribe to outcomes, given her actual information (or the information she ought to have), and ask what choice maximizes expected goodness given those probabilities. The view under consideration—that there are interpersonally objective probabilities, distinct from full-information or morally-ideal-partial-information probabilities, that determine the morally appropriate choice—differs from all of these. But is it plausible that moral assessment concerns *neither* (1) what actually happens, or what idealized (fully informed) actors would choose; *nor* (2) what the actor should choose, given her actual epistemic situation (given her actual probabilistic beliefs, or the beliefs she ought to have given her actual information, or the beliefs she ought to have given the information she ought to have given actual prior information or beliefs)? I doubt it.

Note further that the claimed convergence between what is rational (given OEV) for the Patient to approve, and what is morally appropriate for the uncertain Agent to do in light of the Patient's wel-

---

<sup>45</sup> For a recent attempt to address the problem, see TED LOCKHART, *MORAL UNCERTAINTY AND ITS CONSEQUENCES* (2000).

<sup>46</sup> If we assume indeterminism, this latter view differs from the first.

fare, does not demonstrate that rational approvability has moral weight in a foundational sense. Rational approvability is not an independent and irreducible factor that, among others, determines what the Agent ought morally to do.<sup>47</sup> Rather, what the Agent morally ought to do is determined by (1) the various factors bearing upon the moral goodness of outcomes; and (2) the probabilities that various possible actions will lead to various outcomes. Rational approvability may *track* moral rightness, given OEV, but it does not determine it. Consider, by contrast, the fact that the Patient would suffer terrible pain in some outcome *W*\*. This fact bears upon, and may change, what the Agent ought to do in light of the Patient's welfare, nor can it be reduced to other welfare-relevant considerations.

### C. *Realized Utility and Realized Value*

A realized utility (RU) account of rationality ignores subjective or objective probabilities, and instead looks to actual outcomes: the rational choice is identified, here, as the choice the actual outcome of which maximizes the actor's utility. A realized value (RV) account, too, looks to actual outcomes, but the relevant feature of actual outcomes is their value for the actor, in some sense—e.g., their welfare-value or agency-value—rather than the extent to which the outcomes satisfy her preferences.<sup>48</sup>

Given RU, the rationally approvable action for the Patient is necessarily the action with the best outcome for the Patient's welfare if, and only if, the preferentialist view of welfare is correct. Given RV, assuming the "value" of outcomes is their welfare-value for the actor, the choice that is rationally approvable by the Patient is necessarily the choice with the best outcome for her welfare—regardless of whether the preferentialist view, the substantive view, or some other view of welfare is correct. RV thus creates a tight link between rational approvability and moral weight (since the Patient's welfare does, of course, have moral weight for the Agent); the preferentialist, at least, will also see RU as creating this link.

But RU and RV are, I suggest, highly implausible accounts of rationality. Surely the rational choice, in a given choice situation—or at least some choice situations—depends upon the probabilities with

---

<sup>47</sup> See generally SHELLY KAGAN, *NORMATIVE ETHICS* (1998) (presenting a general template for moral theories that makes moral assessment depend on moral "factors").

<sup>48</sup> See FUMERTON, *supra* note 2, at 102-13 (presenting, but then rejecting, the "actual consequence act consequentialist" account of rational and morally right action).

which different possible outcomes might result from the various actions available to the actor—be they subjective probabilities, objective Bayesian probabilities, objective frequentist probabilities, or probabilities in some other sense. This is the central insight of SEU, whatever its other flaws. RU and RV are detached, not only from the actor's epistemic situation, but from probability of any kind.

#### D. *Other Accounts of Rationality*

The alternative accounts of rationality considered in this Part are all cousins of SEU, in the sense that they all make rationality a function of the outcomes possibly resulting from choice and, with the exception of RU and RV, the subjective or objective probabilities of those outcomes. SEU, OEU, SEV, OEV, RU, and RV, however, hardly exhaust the set of possible or plausible accounts of rationality. Whether rational approvability has moral weight within "plan" accounts of rationality, "deontological" accounts, "expressive" accounts, "minimax" and "maximax" accounts, and yet others that have been advanced within the philosophical literature, is an issue that I cannot pursue here.<sup>49</sup>

#### CONCLUSION

In considering the moral force of ex ante efficiency, it is crucial to distinguish between a person's ex ante preference for some action and the goodness of that action for her welfare or in some other way. An action might maximize someone's subjective expected utility but not lead to the outcome best for her welfare (or best in some other way), since the subjective expected utility calculation depends, in part, on the person's actual beliefs—the beliefs of an actual human being lacking full information. A sophisticated response to this point is that ex ante preference and, more generally, rational approvability possess moral weight independent from the initial goodness of outcomes: that the actor has a moral reason to choose an action rationally approvable by some person even if that action doesn't lead to the outcome best for the person's welfare, or best in some other way; or (even more subtly) that the outcomes resulting from rationally appro-

---

<sup>49</sup> Many of these alternatives to SEU are discussed in *VARIETIES OF PRACTICAL REASONING* (Elijah Millgram ed., 2001).

vable actions are, in the final analysis, the best outcomes for that person.<sup>50</sup>

In this Article, I have attempted (among other things) to rebut the sophisticated response. The rational approvability of some action by some Patient is indeed morally irrelevant for the Agent who might choose that action—morally irrelevant both in the sense that rational approvability lacks the foundational moral weight that would be relevant to a perfectly informed Agent, and in the sense that rational approvability does not indicate to an uncertain Agent which choice, morally, he ought to perform.

Assume my analysis succeeds. Can the proponent of *ex ante* efficiency defend its moral weight in some other way? Perhaps she might concede that *ex ante* preference generally lacks moral force but suggest that the universal convergence of everyone's *ex ante* preference on some action is a singular case in which the Agent now does have a moral reason to perform that action. Is the suggestion persuasive? I will have to leave that issue for another day. Still, my analysis (if cogent) has advanced the debate about *ex ante* efficiency by eliminating one important kind of argument in its favor.

---

<sup>50</sup> This even more subtle line of argument is the "final-preference gambit" that I described and criticized above. See *supra* text accompanying notes 13-16 and notes 41, 44 (explaining objections to that gambit within SEU, OEU, and SEV).